Pedestrian Recognition Using High-definition LIDAR

Kiyosumi Kidono, Takeo Miyasaka, Akihiro Watanabe, Takashi Naito, and Jun Miura

Abstract— Pedestrian detection is one of the key technologies for autonomous driving systems and driving assistance systems. To predict the possibility of a future collision, these systems have to accurately recognize pedestrians as far away as possible. Moreover, the function to detect not only people walking but also people who are standing near the road is also required. This paper proposes a method for recognizing pedestrians by using a high-definition LIDAR. Two novel features are introduced to improve the classification performance. One is the slice feature, which represents the profile of a human body by widths at the different height levels. The other is the distribution of the reflection intensities of points measured on the target. This feature can contribute to the pedestrian identification because each substance has its own unique reflection characteristics in the near-infrared region of the laser beam. Our approach applies a support vector machine (SVM) to train a classifier from these features. The classifier discriminates the clusters of the laser range data that are the pedestrian candidates, generated by pre-processing. A quantitative evaluation in a road environment confirms the effectiveness of the proposed method.

I. INTRODUCTION

Various driving assistance systems, such as collision avoidance and pre-crash safety, have been developed recently to improve the comfort and safety of drivers and cars. Research on autonomous driving has also advanced. It is important for these systems to recognize road environments that contain moving objects, such as other vehicles and pedestrians. In particular, pedestrian recognition is one of the key technologies for decreasing accidents between cars and people, which cause heavy traffic casualties. To estimate a future collision accurately, a system has to recognize pedestrians at a long range.

A horizontally scanning laser scanner, known as a light detection and ranging (LIDAR) system, is often used to detect vehicles and pedestrians. LIDAR has the excellent advantages of high spatial resolution and high range accuracy compared with millimeter wave (MMW) radar, but also the following disadvantages: it does not work robustly in bad weather, such as rain and fog, and its detection range is shorter than MMW radar. However, recent improvements of LIDAR sensitivity have lead to higher performance in bad weather except in fog. Moreover, high-definition LIDAR which can obtain dense range data even in the vertical direction is produced these days. It is widely used as a powerful

Kiyosumi Kidono, Takeo Miyasaka, Akihiro Watanabe and Takashi Naito are with the Road Environment Recognition Lab., Toyota Central R&D Labs., Inc., Nagakute, Aichi 480-1192, Japan (e-mail: {kidono, miyasaka, watanabe, naito}@mosk.tytlabs.co.jp)

Jun Miura is with the Department of Computer Science and Engineering, Toyohashi University of Technology, Toyohashi, Aichi 441-8580, Japan (e-mail: jun.miura@tut.jp)

sensor for autonomous driving. And it is expected to be a future on-vehicle sensor for environmental recognition.

A method for recognizing pedestrians from 3D range data acquired by high-definition LIDAR is presented in this paper. A technique for estimating the ego-motion and tracking moving objects using LIDAR has been developed [1], and it has been confirmed that cars, bicycles and pedestrians can be distinguished based on their size and their velocity. However, a risk assessment [2] for safety driving assistance requires the detection of not only people walking, but also people standing near roads and on sidewalks.

Our goal is to recognize even static pedestrians accurately on the basis of the distribution of 3D range data. Our approach does not require any motion cues. The proposed method divides a measured 3D point cloud into clusters corresponding to the objects in the surroundings. Then the pedestrian candidates are extracted by the size of the clusters [3], [4]. Several features are calculated from the 3D point cloud contained in each candidate, and the classifier distinguishes the pedestrians on the basis of the features. Our approach applies a support vector machine (SVM) to train the classifier. No tracking is considered in this work.

To improve the performance of pedestrian discrimination at a long distance, two novel features are proposed: the slice feature and the distribution of reflection intensities. The slice feature is composed of the widths at different height levels of the human body. This feature can represent a rough profile of the human body from the head to the legs. The latter feature is also effective for distinguishing pedestrians from false positives. The wavelength of the laser beam of LIDAR is in the near-infrared (NIR) region. In the field of spectroscopy [5], it is widely known that NIR lights have different reflection characteristics depending on the materials of the target. So the reflection intensity is considered to be effective for object recognition.

In this paper, a quantitative evaluation is carried out using the range data collected from a moving vehicle in a real road environment. The result demonstrates the effectiveness of the proposed method.

This paper is structured as follows: Section II briefly describes previous related work. Section III shows the specification of LIDAR used in this paper. In Section IV, the details of the proposed method for recognizing pedestrians are presented. Section V contains the experimental results and Section VI concludes the paper.

II. RELATED WORK

Image-based approaches [6]-[13] are very popular in studies of pedestrian recognition. Numerous approaches have

Item	Specifications
Scanning rate	10 scans/s
Horizontal field of view	360°
Horizontal angular resolution	0.23°
Vertical field of view	26.8°
Vertical angular resolution	0.4° (64 lines)
Detection range	40 m for pavement 120 m for cars and foliage
Range accuracy	0.02 m
Wavelength of laser beam	905 nm

Fig. 1.   Sample of 3D range data.

| 10 m | 20 m | 30 m | 40 m | 50 m |

Fig. 2.   Sample of a pedestrian over a range of distances.

been proposed to improve the detection performance and the processing speed. However, little research has been conducted on LIDAR-based pedestrian recognition in 3D range data.

Arras et al. [14] detected people in 2D range data from LIDAR in a cluttered office environment. They used a LIDAR sensor with one horizontally scanning line and applied the AdaBoost algorithm to learn a robust classifier. The classifier, learned from 14 features, such as the number of laser points and the values indicating the linearity and the circularity in the 2D plane, identified the groups of beams that correspond to people's legs. Premebida et al. [15] extended the method to pedestrian detection in a road environment using multilayer LIDAR. They applied a multiple classification method to detect people within a range up to 35 m. Those methods rely on the linearity and the circularity of the 2D range data for the feature extraction. Extending the approach to 3D laser range data increases the computational load.

Spinello et al. [16] expanded these approaches based on a 2D point cloud into a 3D point cloud. They subdivided the 3D point cloud of the target into several 2D point clouds at different heights. The classifiers obtained by AdaBoost estimated whether each 2D point cloud was a part of a human body. The estimation results of all parts of the target were integrated and the target with the correct combination of parts was identified as the pedestrian. However, the detection performance was very sensitive to the distance to the target. For driving assistance systems, it is necessary to improve the performance at a long distance.

Navarro-Serment et al. [17] also presented a method of tracking people in 3D range data from high-definition LIDAR. The 3D point cloud in the target was divided into three parts corresponding to the legs and the trunk of a pedestrian, and the variances of the 3D points contained in each part were utilized as a feature to discriminate the pedestrians. In addition, they represented a 3D pedestrian shape by 2D histograms on two principal planes. The approach has the advantage of a low computational load because the feature extraction is very simple. However, the performance at a long range was also reduced in this approach.

The slice feature and the distribution of reflection intensities are the two novel features proposed in this study to achieve high performance even at a long range.
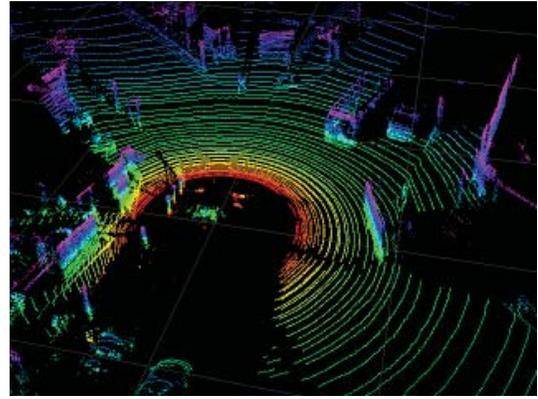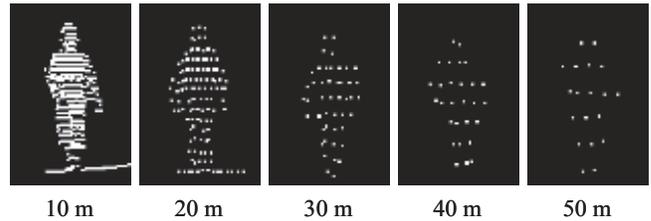
III. HIGH-DEFINITION LIDAR

A Velodyne HDL-64ES2 laser scanner is used as the high-definition LIDAR in this paper. Its specifications are listed in Table I. It has 64 scanning lines aligned at approximately 0.4 degree intervals in the vertical plane and can obtain dense 3D range data by the horizontal scan of 360 degrees. The sensor is mounted on the roof carrier of an experimental vehicle, whose height is approximately 2 m. Fig. 1 shows a sample of measured 3D range data that is composed of points with 3D positions and reflection intensities.

Fig. 2 shows pedestrian samples over a range of distance. Only a few laser beams are irradiated on the pedestrian at 50 m ahead, although the vertical resolution of the sensor is high.

IV. PROPOSED METHOD

The proposed method for pedestrian recognition is described in this section. First, the processing flow from the measurement to the classification is briefly introduced. Then the details of the classification process, with a focus on feature extraction, are explained.

A. Overview of Processing Flow

The processing flow at each scan is shown in Fig. 3. And each process is explained in detail.

1) Data acquisition: A 3D point cloud is acquired from LIDAR.

2) Segmentation: The acquired 3D point cloud is divided into two classes, ground plane and objects, by using an occupancy grid map [18]. All of the 3D points are projected onto the 2D occupancy grid, which is parallel to the ground

(1) Data acquisition   (2) Segmentation

(4) Classification   (3) Clustering
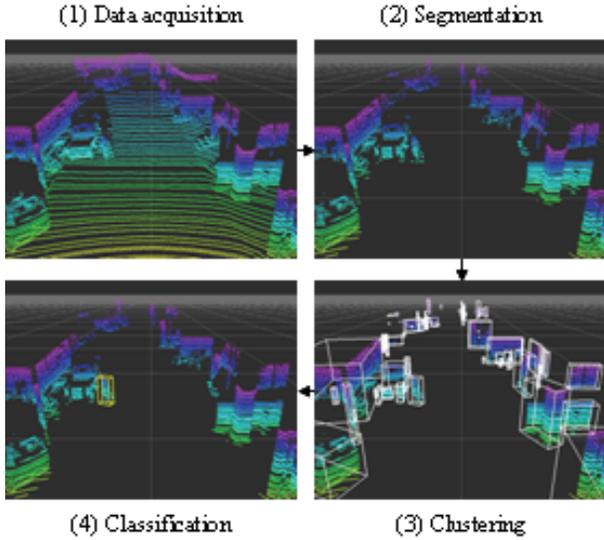
Fig. 3. Procedure of pedestrian recognition.

No.	Description	Dim.[1]
f_1	Number of points included the cluster	1
f_2	The minimum distance to the cluster	1
f_3	3D covariance matrix of the cluster	6
f_4	The normalized moment of inertia tensor	6
f_5	2D covariance matrix in 3 zones, which are the upper half, and the left and right lower halves	9
f_6	The normalized 2D histogram for the main plane. 14×7 bins	98
f_7	The normalized 2D histogram for the secondary plane. 9×5 bins	45
f_8	Slice feature for the cluster	20
f_9	Distribution of the reflection intensity, which is composed of the mean, the standard deviation and the normalized 1D histogram	27

[1]Dim. is Dimension.

plane. The difference between the maximum height and the minimum height in each cell is investigated. If the difference is larger than a threshold, the points in the cell are segmented as objects. The cell size of the occupancy grid is 0.1 m and the threshold of the height for segmentation is 0.3 m.

3) Clustering: The clusters corresponding to vehicles and pedestrians are generated by a distance-based clustering algorithm. To reduce the computational burden, the clustering process is carried out on the occupancy grid by using the labeling technique for image processing. If the distance between two points is within 0.5 m, these points are integrated into the same cluster. A rectangular parallelepiped is applied to each cluster by the Rotating Calipers method [19]. Then the pedestrian candidates are extracted on the basis of the size of the clusters. The conditions are as follows:

$$0.8 \leq h \leq 2.0, \tag{1}$$
$$w \leq 1.2, \tag{2}$$
$$l \leq 1.2, \tag{3}$$

where h, w and l denote the height, the width, and the length of the cluster, respectively.

4) Classification: A feature vector is computed from the 3D point cloud of each candidate and evaluated to classify the candidate into a pedestrian or not. The proposed method applies SVM with a radial basis function (RBF) kernel to learn the classifier.

B. Pedestrian Classification

Table II lists all nine features used in the proposed method. The set of feature values of each candidate C_j forms a vector $\mathbf{f}_j = (f_1, ..., f_9)$. Features f_1 and f_2 are introduced by the Premebida method [15]. The features from f_3 to f_7 are proposed by the Navarro-Serment method [17]. To improve the classification performance, the proposed method adds the following two features.

Fig. 4. The slice feature.

1) Slice Feature for a Cluster: The pedestrian candidates may contain false positives, such as trees and poles, signs, and partially occluded objects. It is necessary to accurately distinguish the false positives from the pedestrians. The pattern of the legs and the profile from the head to the shoulder are distinctive human shapes. It is, however, difficult to extract these partial features at a long distance, where the spatial resolution decreases. A rough profile from the head to the legs is therefore utilized as the 3D shape of the pedestrians.

Three principal axes for the pedestrian candidates are calculated by principal component analysis (PCA). We assume that most pedestrians are in an upright position, so the principal eigenvector is expected to be vertically aligned with the person's body. 3D points in the cluster are divided into N blocks of the same size along the principal eigenvector, as shown in the left image of Fig. 4. As a result of the division, a common feature can be extracted from pedestrians of different heights, such as an adult and a child. Then, the 3D points in each block are projected onto a plane orthogonal to the principal eigenvector, and two widths along the other eigenvectors are computed as the feature. The number of blocks is 10. The feature vector is represented as follows.
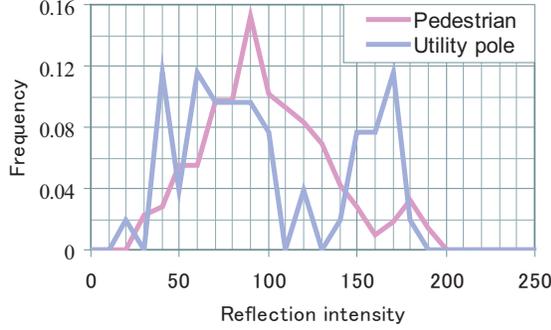
Fig. 5.   Reflection intensity profile of a pedestrian and a utility pole.

$$f_8 = \{w_{10}, w_{11}, \cdots, w_{j0}, w_{j1}, \cdots, w_{N0}, w_{N1}\} \qquad (4)$$

This feature vector is called the "slice feature" in this paper.

2) Distribution of Reflection Intensities in the Cluster:
As known in the field of spectroscopy, a substance has its own unique reflectance characteristics. The reflection intensities of a pedestrian might have large individual variability because their clothing and their belongings are composed of various materials. It is, however, expected that materials leading to false positives, such as trees and utility poles, are comparatively homogeneous. Therefore, the reflection intensities of 3D points can contribute to the discrimination of pedestrians and false positives. Examples of the reflection intensities of a pedestrian and a utility pole are shown in Fig. 5.

The reflection intensity P_r is defined by the following equation. It is in inverse proportion to the square of the distance r:

$$P_r = \frac{P_0 k \sigma}{r^2}, \qquad (5)$$

where P_0 is the intensity of the emitted laser beam and k is the coefficient defined by the LIDAR specifications, and σ is the reflectance of the target.

Strictly speaking, the reflection intensity has to be normalized by the square of the distance. We can directly use the sensor output since LIDAR outputs an 8 bit value normalized for the distance as the reflection intensity.

The following three values are computed from the reflection intensities of the 3D points contained in each candidate:

  i)   Mean intensity
  ii)   Standard deviation of the intensities
  iii)   Normalized histogram: the number of bins is 25 and the range of the intensities is divided at equal intervals.

V. RESULTS

To confirm the effectiveness of the proposed method, the quantitative evaluation is carried out using 3D range data collected in a real road environment by high-definition LIDAR. This section describes the experimental condition and the evaluation results.

TABLE III
CONDITIONS FOR FUNDAMENTAL EVALUATION

(a) Condition of overall evaluation and feature evaluation

Description	Total	N pos.	N neg.
Training data	7700	3700	4000
Evaluating data	8220	4165	4055

(b) Condition of evaluation at different ranges

Description	Total	N pos.	N neg.
Training data	1452	726	726
Evaluating data	1452	726	726

TABLE IV
CONDITION FOR EVALUATION IN ROAD ENVIRONMENT

Description	Total	N pos.	N neg.
Training data	7190	3190	4000
Evaluating data	78977	3190	75787

A. Experimental Condition

Two data sets are prepared for the evaluation. The samples of positives and negatives are manually extracted from the pedestrian candidates, generated by the procedure in Section IV-A.

Data set I contains 11 people who walk around a parked experimental vehicle. The total number of positive samples is 7865. The measurement range is up to 50 m. Five people have some belongings, such as a briefcase and a backpack. Negative samples are extracted from the scenario without pedestrians and their total number is 8055.

Data set II is composed of 75 scenarios in the real road environment. The total number of frames is 16210. The number of positive and negative clusters is 6380 and 79787, respectively. This data set has 124 people with various clothing and belongings. The measurement range is also up to 50 m.

B. Fundamental Evaluation

The fundamental performance of the proposed method is evaluated using data set I.

Fig. 6 shows the classification result for the overall samples within 50 m. This is the receiver operating characteristic (ROC) curve. The vertical and the horizontal axes in the graph represent the true positive rate and the false positive rate, respectively. Table III-(a) shows the condition for training and evaluating. The result of the Navarro-Serment method is shown as a reference. The true positive rate of the proposed method is approximately 0.1 higher than that of the Navarro-Serment method at the point where the false positive rate is 0.01.

The classification ability of each feature is evaluated to confirm the effectiveness of the proposed features. The result is shown in Fig. 7. The two additional features, which are the slice feature and the distribution of the reflection intensities, are more discriminative for the pedestrians than
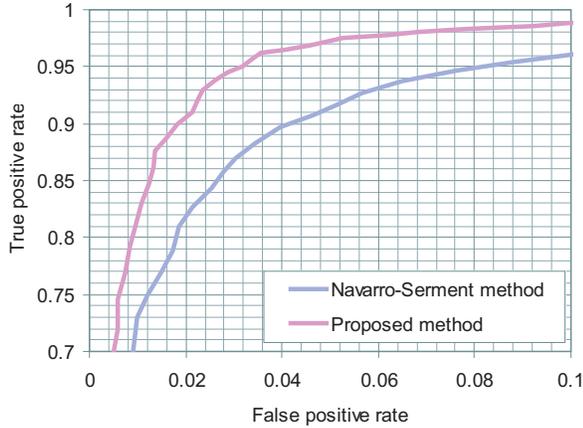
Fig. 6.   Recognition performance for overall samples in data set I.
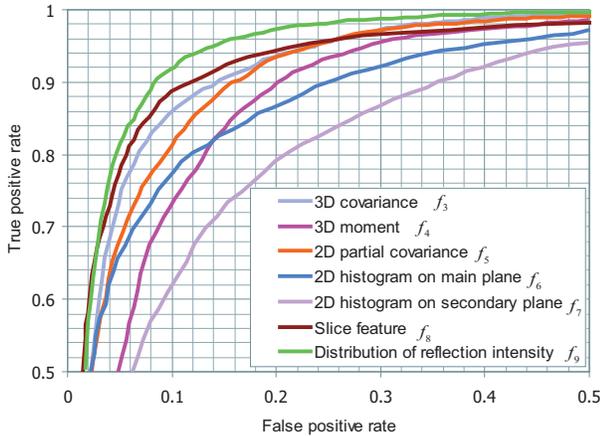
Fig. 7.   Recognition performance for each feature.

(a) 10 - 20 m

(b) 20 - 30 m

(c) 30 - 40 m

(d) 40 - 50 m

Fig. 8.   Results of evaluations at different ranges.

the conventional features. The classification ability of the 2D histogram features for the pedestrian shape is not as high. Note that the slice feature attains a high performance, although it has a small dimension.

To evaluate the performance at different ranges, all samples are divided into four range classes: from 10 to 20 m, from 20 to 30 m, from 30 to 40 m, and from 40 to 50 m. The experimental condition is also listed in Table III-(b). For training and evaluating, the same number of samples are extracted from each range class. Fig. 8 shows the result of the evaluation at different ranges. The proposed method generally shows higher performances than the Navarro-Serment method, especially at the ranges between 30 to 50 m. The two additional features work effectively at the long range, where the spatial resolution of the measurement decreases. It is considered that the shape representation by the 2D histogram is sensitive to the number of 3D points on the targets.

C. Evaluation in Road Environment

The classification performance in dynamic road environments is evaluated using data set II. The numbers of training and evaluating samples are listed in Table IV. Examples of

the recognition results are shown in Fig. 9. The ROC curve is shown in Fig. 10. The horizontal axis indicates the number of false positives per frame. Although data set II contains many women wearing a skirt and putting up a parasol who are not included in the training samples, the high true positive rate of approximately 85% is achieved at 0.1 false positives per frame.

VI. CONCLUSION

This paper presents a method for recognizing pedestrians from 3D range data acquired by high-definition LIDAR. The slice feature and the distribution of the reflection intensities are proposed to improve the recognition performance at a long range with low spatial resolution. The slice feature can represent the rough profile of a pedestrian shape efficiently and the distribution of the reflection intensities is effective to discriminate the materials of the targets. The quantitative evaluation using real 3D range data confirms that the proposed method achieves higher performance than the Navarro-Serment method. Moreover, the proposed features can improve the classification ability at a range of more than 30 m.

However, further improvement of the performance at a long range is required for driving assistance systems. To achieve high accuracy, a fusion system using LIDAR and an image sensor with high spatial resolution will be developed in the future. Other future work is the development of robust
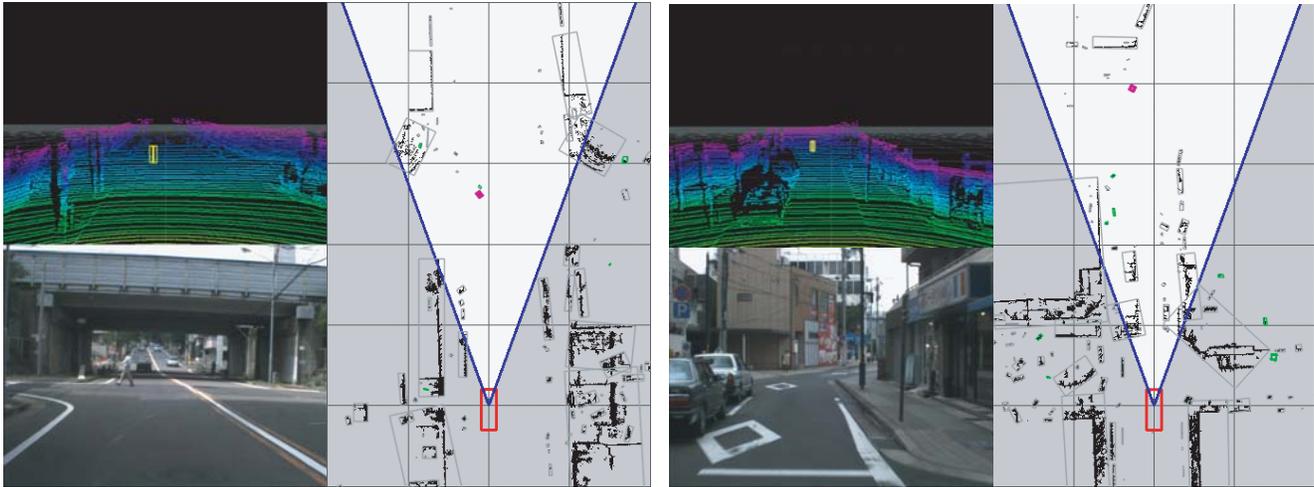
(a) Result for a pedestrian crossing the road.

(b) Result for an oncoming pedestrian.

Fig. 9. Examples of the results of pedestrian recognition in the road environment. For (a) and (b), the upper-left image is the drawn range data as seen from the driver's viewpoint. The yellow box shows the recognized pedestrian. The lower-left image is the video image obtained at the same time as the range data. The right image is the bird's-eye-view image of the range scan. The magenta and green boxes show the recognized pedestrians and the pedestrian candidates extracted by size, respectively. The red box represents the experimental vehicle. The blue lines are the boundary of the camera's field of view. The interval of the grid line is 10 m.
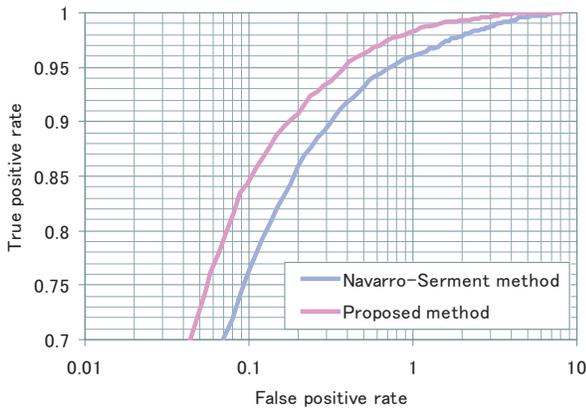
Fig. 10. Recognition performance in dynamic road environments.

features for occluded objects and the identification of other objects, such as bicyclists and people with strollers or carts.

REFERENCES

[1] T. Miyasaka, Y. Ohama and Y. Ninomiya, "Ego-Motion Estimation and Moving Object Tracking using Multi-layer LIDAR," in *Proc. 2009 IEEE Intelligent Vehicles Symposium*, pp. 151-156, 2009.

[2] Y. Ohama, A. Takahashi, M. Kokubun and Takashi Naito, "A Hierarchical Threat Assessment Architecture for Driver Assistance Systems in Urban Areas," in *Proc. 15th World Congress on Intelligent Transport Systems*, pp. 1-6, 2008.

[3] S. Sato, M. Hashimoto, M. Takita, K. Takagi, and T. Ogawa, "Multilayer Lidar-Based Pedestrian Tracking in Urban Environments," in *Proc. IEEE Intelligent Vehicles Symposium*, pp. 849-854, 2010.

[4] G. Gate, A. Breheret and F. Nashashibi, "Centralized Fusion for Fast People Detection in Dense Environment," in *Proc. 2009 IEEE Int. Conf. on Robotics and Automation*, pp. 76-81, 2009.

[5] D. Williams, *Methods of Experimental Physics*, Academic Press, vol. 13, 1976.

[6] C. Papageorgiou, T. Poggio, "A trainable system for object detection," *Int. Journal of Computer Vision*, 38(1), pp. 15-33, 2000.

[7] P. Viola and M. Jones, "Rapid Object Detection using a Boosted Cascade of Simple Features," in *Proc. Conf. Computer Vision and Pattern Recognition*, vol. 1, pp. 511-518, 2001.

[8] N. Dalal and B. Triggs, "Histograms of Oriented Gradients for Human Detection," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 886-893, 2005.

[9] M. Andriluka, S. Roth, and B. Schiele, "People-tracking-by-detection and people-detection-by-tracking," in *Proc. Conf. on Computer Vision and Pattern Recognition*, pp. 1-8, 2008.

[10] T. Kobayashi, N. Otsu, "Image Feature Extraction using Gradient Local Auto-Correlations," in *Proc. 10th European Conf. on Computer Vision*, Part I. LNCS, vol. 5302, pp. 346-358, 2008.

[11] T. Watanabe, S. Ito and K. Yokoi, "Co-occurrence Histograms of Oriented Gradients for Pedestrian Detection," in *Proc. 3rd Pacific Rim Symposium on Advances in Image and Video Technology*, pp. 37-47, 2009.

[12] H. Cao, K. Yamaguchi, T. Naito and Y. Ninomiya, "Pedestrian Recognition Using Second-Order HOG Feature," in *Proc. 9th Asian Conf. on Computer Vision*, pp. 628-634, 2009.

[13] P. Felzenszwalb, D. McAllester, D. Ramanan, "A Discriminatively Trained, Multiscale, Deformable Part Model," in *Proc. Conf. on Computer Vision and Pattern Recognition*, 2008.

[14] K. O. Arras, O. M. Mozos, W. Burgard, "Using Boosted Features for the Detection of People in 2D Range Data," in *Proc. IEEE Int. Conf. on Robotics and Automation*, pp. 3402-3407, 2007.

[15] C. Premebida, O. Ludwig and U. Nunes, "Exploiting LIDAR-based Features on Pedestrian Detection in Urban Scenarios," in *Proc. 12th Int. IEEE Conf. on Intelligent Transportation Systems*, 2009.

[16] L. Spinello, K. O. Arras, R. Triebel, R. Siegwart, "A Layered Approach to People Detection in 3D Range Data," in *Proc. 24th AAAI Conf. on Artificial Intelligence*, pp. 1625-1630, 2010.

[17] L. E. Navarro-Serment, C. Mertz, and M. Hebert, "Pedestrian Detection and Tracking Using Three-Dimensional LADAR Data," in *Proc. Int. Conf. on Field and Service Robotics*, 2009.

[18] S. Thrun, W. Burgard and D. Fox, *Probabilistic Robotics*, The MIT Press, 2005.

[19] G. Toussaint, "Solving geometric problems with the rotating calipers," in *Proc. the IEEE MELECON*, A10.02/1.4, 1983.